Name and brief description of initiative:
Genetic Association Information Network (GAIN), a public-private partnership for conducting whole genome association studies.

Brief description of goals of initiative:
The Genetic Association Information Network (GAIN) is a public-private partnership including the Foundation for the National Institutes of Health, Inc. (FNIH), the National Institutes of Health (NIH), and the private sector. Private partners currently include Abbott Laboratories, Affymetrix, Inc., and Pfizer, Inc. With the sequencing of the human genome, lower cost and more efficient technologies, and the exploration of common patterns of genetic variation through the International Haplotype Map (HapMap) project, we now have the techniques and tools necessary to uncover causal genes that contribute to disease susceptibility. GAIN will support a series of "whole genome association studies" designed to identify specific points of DNA variation associated with the occurrence of a particular common disease. Investigators from existing case-control or trio (parent-offspring) studies have been invited to submit samples and data on roughly 2,000 participants for assay of 300,000-500,000 single nucleotide polymorphisms designed to capture roughly 80% of the common variation in the human genome. Once particular areas of the genome are associated with disease occurrence are identified, it should be possible to zero in on specific genes involved in the disease process much more quickly than would otherwise have been possible.

Principal investigator: N/A, FNIH Project Director David Wholley

Program contact information: David Wholley, 301-594-6343, wholleyd@mail.nih.gov

Website address of initiative: http://www.fnih.org/GAIN/GAIN_home.shtml

Brief description of biomedical informatics and computational biology components and their goals:
FNIH has engaged NCBI to provide technical support for the application submission and review process, data collection and analysis, and controlled access data release procedures for GAIN. NCBI is developing a common database to collect the existing studies in their existing form, convert them to a common structural framework, collect and quality assess the genotype data in a common way, and linking it all together, providing stables accessions to the components, and making it all widely available for the first time. Components will include:

1. Essential OS maintenance functions: primary storage with secondary mirror to ensure reliable access; new content stored in single, managed location; disk management utilities and operative system functions optimized to enhance security
2. Assessment activities: reporting functions; assistance to assemble and/or organize submitted data; technical assessment through use of third-party software if needed

3. Production activities: protocol tagging to convert protocols and manuals into standardized XML-based representation; quality assurance to confirm consistent, valid relationships between each variables' semantic and syntactic attributes and aspects such as operational constraints or ranges for responses; privacy assurance review; quality assessment of genotype data; pre-computes to tabulate and publicly report significant genotype-phenotype relationships

Brief description of resources and tools available for sharing:
GAIN aims to release data as broadly and rapidly as possible in a manner that protects the confidentiality of study participants. Both aims are achieved with a two-tiered controlled/open access model that provides, for broad release of data, scientific review, ethical oversight and investigator accountability for sensitive datasets involving a subject's personal health information, for rapid release through a controlled access process. Immediate, unrestricted open access will be provided to summary results, project study designs, and protocol descriptions.

Open access to the scientific community provides:

- summary measures per genetic marker: allele frequency, genotype frequency
- descriptive information and data context for each study: variables measured, protocols used

Controlled access to authorized investigators provides:

- individually coded phenotypes and associated genotypes
- pre-computed, simple unadjusted genotype/phenotype associations
- tabulations of additional genomic features in notable linkage disequilibrium

Brief description of integrative efforts:
Standard ontologies/terminologies: SNP genotyping data will be provided in conformance with dbSNP, as described at http://www.ncbi.nlm.nih.gov/projects/SNP/
Interactions with other initiatives: The database being developed for GAIN is also anticipated to be used for depositing genotype and phenotype data from the planned whole genome association genotyping of 9,000 participants in the Framingham Heart Study (http://www.nhlbi.nih.gov/new/press/06-02-06.htm) and the NIH-wide Genes and Environment Initiative proposed in the President's FY07 budget (http://www.genome.gov/17516707).

Prepared by T. Manolio 06/23/2006